UNCLASSIFIED

AD **267 779**

Reproduced
by the

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

Fred E. Holdrege, Col USAF
Chief

A. Carp
Technical Director

Personnel Laboratory
Aeronautical Systems Division (AFSC)

NOTICES

DEVELOPMENT OF SELF-REPORT TESTS TO MEASURE PERSONALITY FACTORS IDENTIFIED FROM PEER NOMINATIONS

By
Warren T. Norman

University of Michigan

FOREWORD

The project on development of personality measures is being carried out under Contract AF 41(657)-269 with the University of Michigan, Ann Arbor. Dr. Warren T. Norman, Department of Psychology, University of Michigan, is the Principal Investigator. Dr. Cecil J. Mullins is the Monitor for Personnel Laboratory.

Other members of the project staff during this period were:

Dr. Martha T. Mednick
Mr. James Allison
Mr. Paul Slovic
Mr. Edward Schwartz

This is the second report issued under the contract. The first report was *Problems of Response Contamination in Personality Assessment* by Warren T. Norman (ASD-TN-61-43). = AD - 262 433

The Air Force is indebted to Professor Raymond B. Cattell, University of Illinois, for permission to reproduce items selected from the Objective-Analytic Test Battery for experimental use in this program.

ABSTRACT

An experimental battery of personality tests was constructed as part of a project to develop personality tests appropriate for use in selection of applicants for Air Force officer training. Criteria were peer-nomination ratings previously shown to define personality factors that were predictive of Officer Effectiveness Ratings. Rational selection of testing techniques and item forms was supplemented by information from a series of tryouts with small samples. The battery will be administered to a large sample composed of groups from which reliable peer-rating criteria can be obtained for full cross validation.

TABLE OF CONTENTS

LIST OF TABLES

List of Tables (*Continued*)

DEVELOPMENT OF SELF-REPORT TESTS TO MEASURE PERSONALITY FACTORS IDENTIFIED FROM PEER NOMINATIONS[1]

The problem with which this series of investigations is concerned is the development of effective instruments for the assessment of certain personality characteristics. More specifically, the aim is to develop stimulus materials, administrative procedures, and scoring methods that are insensitive to a variety of response sets and faking tendencies, and to establish empirical validities for the resulting measures against peer nomination rating criteria on five previously established personality variables (Tupes & Christal, 1958).

The problems of fakability and response sets or stylistic response tendencies in the assessment of personality variables were discussed, and a variety of methods that have been proposed for dealing with them were reviewed and evaluated in an earlier report (Norman, 1961). This report will first summarize the results of previous investigations bearing on the definition of the variables, previous methods used to assess them, and their relations to certain criterion measures. The major part of the report is devoted to a series of preliminary studies and their bearing on work in progress.

THE VARIABLES

Tupes & Christal (1958) have reported the results of a series of six personality rating studies. The samples included three groups of Air Force OCS students, one sample of undergraduate college students living in fraternities, and a group of graduate students in clinical psychology. (This last group yielded data for two analyses — one based on peer ratings and one based on ratings by other observers.) All subjects were male and had lived together in small groups for periods of from one week to a year or more.

In each of the studies, members of each group were asked to rate one another on a series of bipolar scales drawn largely from the "personality sphere" set proposed by Cattell (1947). Factor analyses of the six matrices of scale correlations and rotations of each factor matrix to orthogonal simple structure resulted in a clear definition of five interpretable personality factors. Tupes & Christal state that

> . . . the five factors differ only slightly from analysis to analysis. In nearly all cases, the major determiners (variables with loadings above .5) are the same, and in general even the minor determiners (variables with loadings between .3 and .5) are the same (p. 3).

The five factors can perhaps best be described by listing some of the scales mentioned by Tupes & Christal which load highly on each of them.

Factor	Proposed Name	Scales
I	Surgency	Assertive — Submissive
		Frank — Secretive
		Energetic — Languid
		Talkative — Silent
		Adventurous — Cautious
		Sociable — Self-contained

1

Factor	Proposed Name	Scales
II	Agreeableness	Cooperative — Obstructive Attentive — Aloof Goodnatured — Spiteful Mild-mannered — Self-willed Not Jealous — Jealous Emotionally Mature — Demanding
III	Dependability or Conformity	Responsible — Frivolous Conscientious — Unscrupulous Orderly — Indolent Conventional — Eccentric
IV	Emotional Stability	Calm — Emotional Placid — Worrying Poised — Easily Upset Not Neurotic — Neurotic Not Hypochondriacal — Hypochondriacal
V	Culture	Artistic — Not Artistic Cultured — Boorish Imaginative — Practical Polished — Clumsy

The scale names presented in adjectival, bipolar form in the third column above are abbreviated labels and in most cases were expanded or elaborated for actual data collection purposes.

Instructions to raters and the 20 scales (four for each of the five factors) that have been used most generally in subsequent studies are given in the Appendix.

Although rotations to orthogonal simple structure were employed in the analyses by Tupes & Christal, an interesting phenomenon appears when the intercorrelations among the separate scales are examined. Examination of Table 1 reveals that all scales with median loadings of .5 or higher on a common factor have moderate to high intercorrelations, as should be expected. In addition, the correlations between scales which load highly on two different factors are generally quite low. There is, however, one exception. The correlation among scales loading highly on factor II and those loading highly on factor IV have a median value of about .45. This result plus the fact that the correlations between scales defining other pairs of factors also tend more often to be positive than negative suggests that a better representation of the factor structure might be obtained by use of an oblique rotational method. In any event, these data and some to be presented later in this report indicate clearly that factors II and IV are not orthogonal, whereas most of the other factor relationships do approach orthogonal simple structure rather well.

One further study (Wherry et al., 1959) deserves mention and comment relative to our discussion of these trait-rating factors. Wherry and his associates describe a series of rating studies prompted mainly by the Tupes & Christal results and the earlier study by Tupes (1957). Using a 240-item "behavior-gram" checklist which was developed on the basis of the earlier results, they obtained ratings from 100 undergraduates of some "college-age young man who was well known" to the rater. A Wherry-Winer factor analysis yielded a general evaluative factor plus six group factors. The first five group factors obtained correspond very closely to the five factors described by Tupes & Christal. The sixth one was identified as *Having Physical Vim and Speed* and had been predicted to occur as a factor separate from the first group factor (called *Surgent-Extroverted*) since additional items of this sort had been included on the basis of some of the earlier findings (Tupes, 1957).

2

TABLE 1. Intercorrelations[a] and Factor Loadings of the Most Commonly Used Scales

Factor	Scale	Median Loading on Factor[b]	1	2	3	4	5	6	7	8	9	10	11	13	14	15	16	17	18	19
												III				**IV**			**V**	
I	1 Talkative	.78		74	71	66	07	-12	-34	-02	-30	-22	-34	04	00	15	-02	-03	00	04
	2 Frank	.75			72	64	19	03	-24	13	-32	04	-14	30	20	38	13	04	14	18
	3 Adventurous	.72				75	25	14	-16	16	-44	02	-20	37	34	44	27	08	16	20
	4 Sociable	.70					32	06	-13	19	-47	-13	-22	19	24	30	16	08	08	22
II	5 Goodnatured	.77						74	66	72	-29	35	44	48	44	30	48	08	24	34
	6 Not Jealous	.72							72	65	-16	41	51	56	52	36	60	11	32	33
	7 Mild	.68								66	00	42	63	40	36	20	44	12	28	32
	8 Cooperative	.64									10	68	68	54	40	43	58	25	41	48
III	9 Orderly	.64										41	38	00	-14	-01	06	31	22	21
	10 Responsible	.64											72	48	36	52	49	34	49	54
	11 Conscientious	.55												38	22	34	40	34	44	43
	12 Persevering	.50																		
IV	13 Not Neurotic	.71													61	61	68	23	30	40
	14 Placid	.69														68	64	15	30	41
	15 Poised	.68															56	32	51	53
	16 Not Hypochon-driacal	.60																20	35	42
V	17 Artistic	.79																	69	70
	18 Cultured	.75																		74
	19 Polished	.69																		
	20 Imaginative	.68																		

[a] Intercorrelations drawn from Table 2 of Tupes (1957), based on 790 USAF OCS students; scales 12 and 20 not included. Decimal points omitted.

[b] Determined from data presented in Appendix C, Tupes & Christal, (1958).

Wherry and his associates then set up four types of rating forms in addition to the standard peer nomination form and administered them to six male undergraduate college student groups whose members had known one another for at least six months. Each member of each group rated all others in his group, using one or another of the five forms. For each of the groups separate factor analyses were performed (using the Wherry hierarchical method with rotations to simple structure and for similarity of factor profile).

The results showed a high degree of comparability from group to group and also across the different rating methods. A general halo or evaluative factor was again found as it had been in the first analysis. Of special interest is the fact that the rating form which employed a strictly forced-choice format yielded loadings on this factor that were considerably lower than those for any other form across all of the content areas. This is especially important relative to some of the test devices we shall describe later in this report.

The second factor represents a fusion of the two factors (or item content areas) A and F (*Surgent-Extroverted* and *Having Physical Vim and Speed*) which had been identified in the previous analysis. This is of particular interest since this factor now appears to be identical with the first factor identified by Tupes & Christal. It was from the content of the Tupes & Christal factor I that the items for the A and F content areas of Wherry *et al.* had been drawn on the assumption that this factor was really heterogeneous in composition.

Factors II, III, and V of the Tupes & Christal analyses were also found in each of the separate group analyses by Wherry and his associates, but no factor corresponding to number IV of the earlier studies was identified. All in all, however, the results of this series of studies must be interpreted as confirming rather dramatically the results of the Tupes & Christal analyses — especially when one considers the differences in subject populations, rating forms, and analysis methods employed.

The finding by Tupes & Christal and the confirmation by Wherry and his associates that the factor structure of peer nomination ratings on a broad set of personal attributes displays marked stability under diverse conditions, across varied subject populations, and using different rating forms is itself of considerable interest and importance. There is, however, an additional finding (Tupes, 1957) of at least some practical importance. When ratings on scales of this sort were correlated with subsequent Officer Effectiveness Reports, moderate to low validities were obtained for some of the scales against the Overall Effectiveness Ratings by superiors. What is more, the magnitude of the validities for scales which later were found to load highly on a given factor were found to be quite consistent. Those scales loading on factors III and V had zero-order validities between .26 and .29. Those loading on factors II and IV were slightly lower on the average and somewhat more varied although all were positive. The four scales which loaded most highly on factor I, however, all had low *negative* correlations with this criterion of rated officer effectiveness. A multiple correlation of .52 is reported by Tupes, based on a sample of 615 of the 790 cases in the original group.

Thus there is some evidence that even within so highly selected a sample as OCS graduates, ratings on certain personality attributes account for an appreciable amount of the variance in rated officer effectiveness.

If measures of these characteristics could be obtained on applicants to OCS prior to admission, such information presumably could lead to the selection of better officer material. It is at this point, however, that a limitation of the peer-nomination rating method becomes critical. In order to place any degree of confidence in ratings of this sort, the participants must have had an opportunity to observe one another in a variety of situations over a period of some time prior to making their ratings. There is evidence in the studies reported by Tupes & Christal that for clinical psychology graduate students, a week of intensive, varied, and intimate association is sufficient and that for groups of OCS students, usable ratings can be obtained after as little as

three weeks of close and continuous association. In general, however, a longer period is usually recommended, and Cattell (1957, p. 63), for instance, argues for a minimal period of no less than two or three months and preferably a year or more.

This feature of the peer-nomination rating method makes it practically unusable in most selection and classification programs. In only the most unusual situations are the applicants for such programs familiar enough with each other on the basis of previous contacts to qualify as sophisticated and informed raters.

Clearly, what is required, if characteristics such as those described above are to be used for selection or classification purposes, is a set of measures of these attributes which may be obtained for each applicant *based only on his own behavior.* If assessment instruments of this sort can be found or developed which produce scores for each subject which relate closely to scores derived from the peer-nomination method, and which are insensitive to contamination and distortion, then useful and effective means will be available for incorporating data on these kinds of personal attributes into selection and classification batteries.

Although the need for such assessment methods is crucial in the context of selection and classification programs, one could argue that there is also a considerable need for such devices in the context of theoretical studies of personality and in diagnostic settings as well. It is in fact only in the most unusual experimental and applied situations where one has available a group of persons who are acquainted intimately enough with a given person to serve adequately as raters of him. Even when such persons are available, this method of assessment is at best uneconomical. Since personality variables of the sort we are presently concerned with are likely to be of theoretical and practical interest in situations other than just the particular one currently being considered, valid and fake-proof devices for assessing such attributes in an efficient and economical fashion should prove to be of considerable general usefulness.

There exist, of course, a large number of inventories and other assessment devices which purport to measure traits more or less like the attributes described above and which produce scores based only on the individual respondent's behavior. By and large, however, they suffer from two important defects. In the first place, most of these devices have not been subjected to careful empirical validation against relevant external criteria. Secondly, there is ample evidence that many of these instruments (especially the self-report questionnaires) are sensitive to faking tendencies and other forms of response distortion and contamination.

This second point is extremely important in the selection context. Persons applying for admission to some program ordinarily want to be selected, and it must be presumed that they will do whatever they can to achieve this end. Although some applicants will cooperate with instructions that ask them to be "as frank and honest as possible," it would be naive to assume that all will do so or that when in doubt about a given answer even the most well-intentioned applicant will respond in a way he believes will present himself in a bad light.

For selection purposes, assessment methods are required which effectively preclude the possibility of response distortion by the examinee and which possess demonstrable validity against relevant external criteria. For this project, the external criteria are available in the 20 scales (Appendix) selected from those established as defining five personality factors. The rationale governing construction of a self-report battery to measure the same personality factors was developed in an earlier report (Norman, 1961).

TEST DEVELOPMENT

The remainder of this report is organized in terms of the development and preliminary standardization and validation of the scales of a variety of assessment devices. Some of these

instruments have been built by the project staff expressly for the purpose of tapping one or more of the five rating dimensions; others have been adapted from previously published tests and inventories; and still others have been employed in their original form in one or more of the development studies. In the course of constructing and standardizing some of these instruments, various kinds of empirical data have been collected from a number of different groups of subjects and have been utilized in the development of stimulus materials, in the standardization of administrative procedures, and in the construction of preliminary scoring keys. A brief description of each of the studies conducted will first be given, including the nature and size of the sample, the task and stimulus materials employed, and the nature of the data collected and their intended use. Then the construction of each of the several tests developed by the project staff will be described, citing where and how data from the various studies were utilized. Following this is an annotated listing of other tests that have been used in this phase of the research program. Finally, the results of analyses obtained from the preliminary validation studies will be presented and evaluated.

SUMMARY OF DEVELOPMENT STUDIES

Study 1: OCS-Desirability Scaling of Personality Descriptive Adjectives and Occupational Titles

Sample. 21 male and 26 female students in a laboratory course in tests and measurements and individual differences at the University of Michigan, Fall semester, 1959.

Task and Materials. Each subject was instructed to rate 193 personality descriptive adjectives and 164 occupational titles in terms of his own Air Force Officer Candidate Desirability stereotype. The adjectives and the occupational titles were presented to the subjects in separate booklets in which each adjective or title was followed by a 9-interval graphic rating scale. The odd-numbered intervals were labeled Very Undesirable, Moderately Undesirable, Neutral, Moderately Desirable, and Very Desirable. The instructions asked the subject to judge how desirable he thought it would be for a man who wanted to become an Air Force officer to have each of the characteristics (adjectives) or occupational preferences (titles).

The Data and Their Use. The rating distributions for each adjective and occupational title were tabulated separately for males and females and means and variances for each distribution were computed. Correlations between mean scale values for the males and those for the females were computed for the adjectives and for the occupations separately to determine the degree of comparability of the data from the two sex groups for each kind of stimuli. The means and variances of rating distributions based on the male subsample were subsequently used to construct forced-choice items for two multiscale inventories — the Descriptive Adjective Inventory (DAI) and the Occupational Preference Inventory (OPI) to be described in more detail below.

Study 2: Normative and Fakability Study of the DAI and OPI.

Sample. 418 Air Force student officers and air cadets, February, 1960.

Task and Materials. The forced-choice Descriptive Adjective Inventory (DAI) and Occupational Preference Inventory (OPI) were administered under the following conditions to the subjects:

> *Group 1* (N = 78) took both instruments under ordinary self-report instructions.
>
> *Group 2* (N = 136) took both instruments under instructions to fake their responses as best they could to gain admission to the Air Force Officer Candidate School
>
> *Group 3* (N = 204) took both instruments, first under standard self-report instructions (straight-take) and subsequently under the admission-to-OCS faking instructions (fake-take).

6

The Data and Their Use. Each subject's responses to all items in both inventories were punched on IBM cards for analysis purposes. Percentage endorsement indexes for each response alternative for each forced-choice item for each sample under each instruction set were computed. The data from Groups 2 and 3 were collected to determine the sensitivity of each item to faking tendencies and the resistance to such tendencies for keys that might subsequently be built for these tests. The data from Group 1 and from the straight-take administration to Group 3 were intended to be used to develop norms and to determine other distribution properties for scoring keys for these instruments. The data from Group 3 are being used (together with other itemetric data) to build sets of preliminary keys for these inventories. Hence, to *test* the insensitivity of such keys to faking tendencies strictly speaking requires that data from independent samples (Groups 1 and 2) be used. A final objective was to obtain data on these devices for comparing the responses of military samples of this sort with the responses obtained from civilian student groups on whom most of the test development research is being done.

Study 3: OCS-Desirability Scaling of Self-Report Statements

Sample. 125 Air Force student officers and air cadets, August, 1960.

Task and Materials. Each of the 1606 personality descriptive statements contained in the four forms of the Self-Report Item Pool (SRIP forms A, B, C, and D) were rated in terms of their admission-to-OCS desirability properties. Five-point rating scales were employed, with the points labeled Very Undesirable, Somewhat Undesirable, Uncertain, Somewhat Desirable, and Very Desirable. The statements in forms A and B (402 and 400 items respectively) were rated by 58 of the subjects and those in forms C and D (403 and 401, respectively) were rated by the remaining 67 subjects. The instructions asked the subject how desirable he thought it would be for someone applying for admission to the Air Force's Officer Candidate School to say "True" to that statement.

The Data and Their Use. All ratings were punched on IBM cards and the means and standard deviations for each item were computed. In addition, correlations were computed among all pairs of items within blocks of 60 items on each form (terminal blocks of 42, 40, 43, and 41 items). These values were used, together with other data for these items, to construct forced-choice, paired statement items for inclusion in the Forced-Choice Self-Report Inventory (FCSRI).

Study 4: Pretest of Cattell's 18 Objective Analytic Battery (Group Form)

Sample. 23 paid male student volunteers, University of Michigan, Fall semester, 1959.

Task and Materials. The major part of the 18 O-A Battery was administered to the subjects in two 3-hour group testing sessions under the standard instructions and time limits specified in the test *Handbook.* The intent of this study was to become familiar with the complicated administrative and scoring procedures for the tests in this battery and to generate rough norms for the variables scored on these tests.

The Data and Their Use. The tests were scored in the established manner and distributions on each variable were tabulated. These data and some of the problems encountered in giving and scoring the tests were evaluated with a view toward clarifying the instructions, modifying the response formats for certain tests, and for making decisions concerning which of the subtests to use in later studies.

Study 5: The ROTC Preliminary Validation Study

    Sample. 84 paid, male volunteers from the senior classes of the three ROTC units at the University of Michigan, Fall semester, 1959.

    Task and Materials. The assessment devices administered included (1) the set of 20 peer-nomination rating scales for assessing the five factors described earlier in this report, (2) the DAI, (3) the OPI, and (4) the Welsh Figure Preference Test (WFPT). Six rating groups were formed, each composed of men within the same ROTC unit, and ranging in size from 12 to 16 men. Approximately one-third of the men in each group were nominated as high and one-third as low on each of the 20 scales by each member of the group, excluding self. Each man then completed each of the three other tests according to standard self-report instructions.

    The Data and Their Use. Scores derived from the peer ratings were computed. A factor analysis of these scores was then carried out to determine the comparability of this sample to those used in the earlier studies. The DAI and OPI were scored on five *a priori* keys based on judgments by the project staff of the relevance of responses to each item for one or another of the peer-rating factors. The WFPT was scored on the eight scales for this test (DL, RP, CF, BW, RA, MF, NP, and MV) which seemed most promising as correlates of the peer-rating factors on the basis of the arguments and data presented in the manual. These test performances have also been scored on other keys built subsequently for these devices.

Study 6: The Fraternity Preliminary Validation Study

    Sample. 82 paid male volunteers from 8 residence groups (6 social fraternities, 1 professional fraternity, and 1 cooperative housing unit) at the University of Michigan. The subjects were predominantly seniors and were recruited and tested during the Spring semester, 1960.

    Task and Materials. Nine rating groups were formed (two from one of the large fraternities), ranging in size from 7 to 11 men. Each man completed 14 hours of tests, inventories, and ratings. The instruments included in the battery were:

1. The standard 20-scale peer rating forms for the five factors
2. A peer-rating scale on risk-taking
3. The USAF Life Experience Inventory (on risk taking)
4. The Descriptive Adjective Inventory
5. The Occupational Preference Inventory
6. The Welsh Figure Preference Test
7. The 16 PF Questionnaire — Form A
8. The 16 PF Questionnaire — Form B
9. The Bet Preference Test (yielding measures of variance and skewness tendencies in betting choices)
10. Self-Crediting Test — V
11. Word Meanings (Standard and Penalty conditions)
12. Word Construction — A
13. Culture — E
14. Dot Estimation
15. Verbal Intelligence (risk taking)
16. General Knowledge — A (adapted from Cooperative General Culture Test — Form X)
17. Part of the Cattell 18 O-A Battery (Tests G2, G6, G8, G9, G10, G11, G13, G15, G16, G17, G18, G19, G22, G23, G24, G27, G30, G32, G34, G35, G37, G38, G41, G42, G43, G44 a and b, G45, G47, G49, G50)

Data were collected on items 1-3 during separate one-hour sessions with each rating group. Items 4-8 were completed by each man at his convenience under self-administration conditions and the remaining instruments, 9-17, were given in three group-testing sessions of about three hours each, separated by one-week intervals. In addition to the above, 55 of the original 82 subjects also completed an additional self-administering 4-hour battery containing the following:

18. An 8-item forced-choice questionnaire called the Job Preference Inventory
19. SRIP, Forms A, B, C, and D

All tests were administered under standard instructions.

The Data and Their Use. Rating scores for each subject on each of the peer-nomination scales were computed and a factor analysis was performed to determine the comparability of this sample with those used in previous studies. These data and performances on the rest of the tests included in the battery were used in a variety of ways for key construction, norm development, and validation purposes.

CONSTRUCTION OF TEST MATERIALS

A number of assessment devices have been constructed for the explicit purpose of developing self-report predictor scales to tap the five peer-rating factors. In the construction of these test forms and scoring keys an attempt has been made (1) to minimize the possible influence of desirability faking tendencies by the respondents, (2) to maximize the empirical validities of the scoring keys against the peer-nomination dimensions, and (3) since the rating factors appear to be relatively orthogonal, to minimize the correlations among the scales for each instrument. Before proceeding to an account of the construction of these devices, some research findings based on the peer-rating scales themselves should be mentioned, since the data collected by use of these scales play a critical role in the development of the other instruments.

THE PEER NOMINATION CRITERION RATING SCALES

The development of these scales and the reasons for selecting the particular ones used in this program of research have been presented in the first part of this report. In brief, the 20 scales are those which, on the basis of previous analyses, seem to best define the recurrent set of five factors.

In each of the two preliminary validation studies (studies 5 and 6 above), the rating scores[2] obtained have been factor analyzed to determine whether the same five factors would emerge and to determine the degree to which the loadings for the several scales would correspond to those found in prior investigations. One of the principal reasons for running the ROTC study was to see whether groups of persons who had not had an opportunity to live together, but who had shared classroom and drill experiences over a considerable time period, would be able to rate each other effectively on these scales. The results of these analyses on the two samples are presented in Table 2.

---

[2] Scores were computed on each scale for each subject according to the formula

$$\text{Rating Scale Score} = 10 + \frac{10}{N-1}(X_A - X_B)$$

where N = the number of persons in the rating group, X_A = the number of times the subject was rated as "A" on the scale by the other members of his group, and X_B = the number of times he was rated as "B" on the scale. The possible range of scores for any size group is thus 0 to 20 with a mean of 10 for all groups.

TABLE 2. Loadings and Communalities of the Criterion Rating Scales

(Based on Data from the ROTC and Fraternity Validity Studies)

ROTC Study

Scale	I	II	Factor III	IV	V	h²
1	.915	-.010	.136		-.008	.856
2	.877	.051	.121		.068	.791
3	.859	.184	-.075		.018	.778
4	.911	.033	-.150		.048	.856
5	.121	.728	.371		.238	.738
6	.017	.818	.298		.038	.760
7	-.456	.630	.156		.335	.742
8	.287	.595	.398		.320	.698
9	-.347	-.066	.735		.247	.726
10	.111	.375	.851		.049	.879
11	-.445	.391	.618		.283	.813
12	.131	.299	.887		.055	.897
13	.344	.629	.487		-.050	.753
14	.540	.512	.025		-.456	.763
15	.606	.412	.384		-.200	.725
16	.578	.542	.185		-.281	.742
17	-.242	.139	.376		.675	.674
18	.179	.209	.537		.562	.680
19	.285	.480	.460		.376	.664
20	.259	.160	.034		.609	.465
"Purity" Index	.61	.51	.61		.63	

Fraternity Study

Scale	I	II	Factor III	IV	V	h²
1	.791	-.178	-.070	-.166	.259	.757
2	.722	.069	.144	-.014	.108	.559
3	.704	.178	-.381	.240	.128	.746
4	.717	.154	-.254	-.095	.122	.626
5	.290	.820	-.033	.166	.151	.809
6	-.048	.784	.071	.372	-.060	.765
7	-.154	.871	.132	.056	-.017	.804
8	.452	.704	.217	.048	.091	.758
9	-.247	-.384	.619	-.301	-.004	.682
10	-.030	.235	.894	-.044	.081	.864
11	-.124	.463	.645	-.107	.196	.696
12	-.074	-.012	.886	.159	.015	.817
13	-.182	.551	-.007	.639	-.085	.752
14	-.028	.404	-.202	.754	-.127	.790
15	-.015	-.037	.187	.736	.154	.602
16	.125	.205	.036	.604	-.223	.473
17	.075	-.126	.226	-.315	.649	.594
18	.164	-.005	.115	-.022	.857	.774
19	.070	.145	.315	.051	.650	.550
20	.186	.098	-.153	.005	.855	.799
	.81	.70	.81	.79	.89	

Factors were extracted by the principal axes method and were rotated analytically to orthogonal simple structure by Kaiser's Varimax method. In Table 2 the scales are grouped in terms of the Tupes & Christal factors they were intended to measure: scales 1-4 for factor I, scales 5-8 for factor II, etc. In addition, the factors obtained in each analysis have been ordered to correspond to the Air Force factor numbers they seem to match rather than the order in which they were extracted. Finally, certain of the obtained factors have been reflected to correspond with the positive poles of each of the Air Force factors.

It is clear from an inspection of the right side of Table 2 that the results obtained from the fraternity sample resemble very closely those from the several studies reported by Tupes & Christal. Only in the case of factors II and IV is there any evidence of nonindependence among the factors — a finding which confirms the earlier results.

The data from the ROTC study, however, yielded only four factors instead of the usual five, even though the same criterion of 98% of variance extracted was used in both analyses. The scales intended to tap factor IV are seen to load consistently and moderately high on both factors II and I. The hyperplane counts are distinctly inferior for these data compared to those from the fraternity groups, and the number of scales with high loadings on irrelevant factors is considerably larger.

These observations are more clearly indicated by the "purity" indexes at the bottom of each column. To obtain these values, the sum of the squared loadings for the four scales intended to tap the factor in that column was divided by the total sum of squared loadings in the column. As can be seen, the values are distinctly lower for the ROTC sample than for the fraternity groups. This, together with the fact that the communalities are slightly *higher* in the ROTC study (even though one less factor was extracted), indicates that some halo-like contaminant is present in the ratings obtained from the ROTC groups.

This impression is further confirmed by an examination of the correlations between the factor scores[3] for subjects in the two studies. The correlations among these rating-factor scores for the two samples are presented in Table 3.

TABLE 3. Correlations Among the Rating-Factor Scores

Factor	ROTC Study				Fraternity Study			
	I	II	III	IV	I	II	III	IV
I								
II	.10				.19			
III	-.10	.63			-.28	.15		
IV	.58	.45	.26		-.03	.46	-.06	
V	.13	.56	.53	.28	.25	.10	.22	-.16

On the basis of the prior studies reported by Tupes (1957) and Tupes & Christal (1958), only the magnitude of the (II, IV) Correlations could be expected to depart appreciably from zero if these groups had been comparable to those used earlier. In the ROTC group, however, four of the other correlations exceed the .45 value obtained (and expected) between factors II and IV. The correlations for the fraternity sample, however, are much more in line with the previous

---

[3] The five factor scores for each subject were computed by simply summing his scores on the four rating scales representing each factor in the set of 20 scales.

findings, and indicate that, except for the moderate positive relation between factors II and IV, an essentially orthogonal structure exists among these variables for this group.

However, by no means can all of the variance in the ratings for the ROTC sample be attributed to a generalized rating stereotype. In the first place, four factors emerged in the factor analysis of these data and it was not too difficult to identify which of the five previously reported factors each of these corresponded to. Also, from an inspection of Table 3, it is clear that the scores for factor I are clearly independent of those for all but factor IV, and those for factor IV are essentially independent of the scores on factors III and V.

In a final attempt to determine just how much correspondence exists between the two sets of factors derived from these two samples, the four factors that emerged in the ROTC analysis were related to the set of five factors from the Fraternity Study analysis by a method recently developed by Kaiser (1960). This method overcomes some of the interpretational difficulties that characterize other methods of factor "matching" and yields a matrix of "relational" indexes which are interpretable as ordinary product-moment correlation coefficients. The transformation matrix of relational indexes is given in Table 4.

TABLE 4. Relationship Indexes Between Factors From the ROTC and Fraternity Studies

		Fraternity Study Factors				
		I	II	III	IV	V
ROTC	I	(.88)	-.07	-.17	.40	.18
Study	II	-.12	(.91)	-.03	.38	.06
Factors	III	.06	-.05	(.96)	.19	.16
	V	.11	.20	-.02	-.57	(.79)

The table has been organized again in terms of the Air Force factor labels, i.e., certain factors have been reflected and reordered to correspond to the earlier labeling. The values enclosed in parentheses are the indexes between corresponding factors as designated in Table 2. These values are quite high as correlational values generally run, but it should be borne in mind that these are correlations between *factors* rather than between measure variables and may require a different standard for interpretation. Despite these seemingly high indexes, it was felt that the evidence for the operation of a halo-type contaminant in the data from the ROTC sample as presented in Tables 2 and 3 was sufficiently clear to contra-indicate the use of the ROTC data either for test validation or for designating criterion groups for empirical scale construction. It was also decided on this basis to limit all subsequent validational studies to groups whose members had had an opportunity to associate with each other in the more intimate context of residence settings.

As a sidelight to these analyses of the criterion ratings, the factor which failed to appear in the ROTC analysis (factor IV) is the same one which failed to appear in the recent studies by Wherry *et al.* (1959) who also used some groups of AFROTC students who were not living together.

In the Fraternity Study, the scores on each factor ranged over practically the entire possible range (0 to 80) in a roughly continuous fashion. The distributions of factor scores were symmetric and platykurtic for all factors, as one would wish them to be to obtain maximal interpersonal discrimination. These facts indicate that despite the forcing of ratings within groups (or perhaps because of it) with the consequent equating of mean scores for all groups, there was a considerable

amount of agreement on the part of the raters in making their nominations. If it were not so, the distributions should have displayed a more nearly normal (random error distribution) form with an attendant restriction in range (since there were only 82 cases in the sample).

Our general evaluation of these scales, then, is that they yield interpretable, sensitively discriminating measures on a set of relatively independent personality attributes when they are employed in groups with appropriate associational backgrounds.

THE DESCRIPTIVE ADJECTIVE INVENTORY (DAI)

When considering the choice of stimulus materials with which to construct tests for these personality dimensions, trait descriptive adjectives come to mind at once. The peer rating factors are, after all, derived from scales whose extremities are defined for the raters by means of just such terms. On the other hand, we have been at pains to stress the pervasive and distorting influence of general evaluative response sets in the use of self-descriptive materials for which attribute reference and evaluative valence are easily discernible by the respondent (Norman, 1961). These arguments, generally presented in criticizing the use of self-report statements, are probably even more cogent in the case of simpler, less-likely-to-be-ambiguous, one-word predicates.

However, it seems reasonable that the primary difficulty lies, not in any inherent deficiencies of these kinds of stimuli but rather in the way in which such test stimuli typically are presented to the assessee — i.e., in a relatively free-response format such as a checklist or True-False inventory. If this is so, and if the instructions to the respondent and the response format available to him can be modified so as to eliminate (or at least markedly reduce) the effects of such influences, then a valuable source of relevant, easily used test materials immediately becomes available. In developing the Descriptive Adjective Inventory and two other self-report tests, it has been assumed that this diagnosis of the problem is correct and that the remedy consists in the sophisticated use of such materials rather than in their categorical rejection.

The adjectives chosen for use in this instrument were drawn primarily from a pool of 342 trait descriptive terms that had been used by Dunnette[4] in an earlier item scaling study. This list of 342 adjectives was sorted by several members of the research staff into factor categories. The 193 items for which there was high interjudge agreement as to factor-pole relevance were compiled into a booklet form — The Adjective Rating Schedule — for purposes of collecting "Admission-to-OCS-desirability" ratings for each term.

The subjects and rating procedures used in obtaining single-stimulus desirability data on these adjectives and the analysis and use of these data to construct binary forced choice items are presented above in the description of Study 1. The data obtained from the males and females in the sample were analyzed separately. The mean OCS-desirabilities correlated .98 between the two samples, which is in line with similar results for self-report statements reported by Edwards (1954). The data from the male subsample alone were used as a basis for item pairing, however, since there was evidence of a sex difference in another rating task by this sample and because the inventory to be built was intended for use, initially at least, with males.

A mean-by-variance scatterplot of the adjectives was constructed and binary items were formed by matching terms representing two different factors (based on staff judgments of their content) as closely as possible on these two desirability parameters. Actually some difference in means and variances for paired stems was permitted, although no two were paired for which the mean difference was greater than .3 scale units. The number of times any given adjective was paired with others varied from once to four times depending on the location of the term and the density of other factor representatives in its vicinity. The distribution of points in the

---

[4] Dr. Marvin D. Dunnette, reported in a personal communication to the author, December 1959.

scatterplot (means along the abscissa and variances along the ordinate) had the usual kidney shaped outline — low variances for the large number of adjectives near the extremes of the abscissa, with generally larger variances for the smaller number with mean desirabilities near the middle of the scale.

A few items were formed by pairing adjectives which had been rated with ones drawn from Roget's Thesaurus which had not been included in the rating study. This was done to obtain a better balance in the inventory of items judged on *a priori* grounds to be reflective of the several factor combinations. In all, 200 binary items were constructed. These were arranged in a roughly systematic fashion which attempted to alternate factor combinations, A-B positions, and plus-minus valenced pairs in cycles of about 20 items throughout the inventory. This was done to break up any short span serial position effects which might otherwise operate. The forced-choice items thus constructed and arranged were typed in booklet form for use with a separate answer sheet and the test was named the *Descriptive Adjective Inventory* (DAI).

DAI a priori keys. A priori keys for each of the five factors were constructed on the basis of the staff judgments of each adjective's factor relevance. Every adjective in the inventory was scored on one and only one of these keys. Thus relationships among those keys contain no multiple-scoring artifacts, but do have a built-in negative bias owing to the forced-choice nature of the responses and the exhaustive scoring of all response categories. The extent of the bias in these interkey relationships and the relationships between these keys and the criterion rating factor scores as estimated on the ROTC and Fraternity Validity Study samples can be seen in Table 5.

TABLE 5. Correlations Between the DAI A *Priori* Keys and the Criterion Rating Factor Scores

DAI A *Priori* Keys	Factor Scores					DAI A *Priori* Keys				
	I	II	III	IV	V	I	II	III	IV	V
ROTC Sample (N=84)										
I	(.50)	-.12	-.35	.23	-.07					
II	-.19	(.24)	.16	-.05	.05	-.41				
III	-.25	.26	(.37)	-.09	.00	-.42	.04			
IV	-.19	-.07	.04	(-.01)	-.03	-.21	-.31	-.07		
V	.11	-.21	-.16	-.11	(.16)	.23	-.31	-.54	-.31	
Fraternity Sample (N=82)										
I	(.32)	.02	-.11	.02	-.09					
II	.05	(.23)	-.17	.04	-.13	-.32				
III	-.37	-.15	(.40)	-.15	-.29	-.43	-.13			
IV	-.07	-.01	.01	(.26)	.17	-.18	-.24	-.16		
V	.14	-.06	-.21	-.08	(.40)	-.01	-.28	-.51	-.20	

Note. — Numbers in parentheses are validity coefficients; a correlation of .21 is significant at the .05 level, of .28 at the .01 level.

The only possible exceptions to the otherwise uniform pattern of negative interscale correlation occurs between scales I and V and scales II and III in the ROTC sample and only the (I, V) correlation is appreciably different between the groups.

The relationships between these keys and the criterion rating scores are, however, considerably less uniform. We have already presented the basis for our distrust of the rating data from the ROTC sample — especially those from scales defining factor IV — and Table 5

14

provides additional reasons for this evaluation. These subjects' responses to items judged to be reflective of factor IV are unrelated to their scores on this factor derived from peer ratings. However, excluding the entries for factor IV, in only two instances is a heterotrait-heteromethod value larger than the two corresponding validity entries in this sample. Turning to the data from the fraternity sample one observes a more uniform set of validities across all five factors and only one exception to the principal criterion for discriminant validity – the -.37 value for the factor III DAI scale against factor I ratings, the magnitude of which slightly exceeds the validity coefficient for factor I (but not that for factor III).

Considering the facts (1) that the rating factor scores are generally uncorrelated (but *positively* related in the ROTC sample owing to halo effects), (2) that a built-in negative correlation exists between the *a priori* scales of the DAI, and (3) that these scales are based only on content judgments of the relevance of individual adjectives to the rating factors (taking no account of the effects of the forced choice context on the respondents' choices nor of any empirical item validity data), these results were considered quite encouraging.

DAI preliminary empirical keys. The next step was to build a set of scoring keys which would not be subject to these shortcomings. Such keys should be constructed so as to maximize the validity of each against the corresponding rating factor, to be mutually uncorrelated (with the exception of the keys for factors II and IV which optimally should have a moderate, positive correlation to match that between these two factors in the rating domain), and finally, to be insensitive to faking tendencies. Hence item analyses were performed using primarily the data from the Fraternity Validity study and from the OCS-desirability faking study (Study 2). The procedure followed in the construction of these preliminary empirical keys was:

1. Split the distribution of rating scores for the fraternity sample on each factor at the median.

2. For the first alternative of each item, calculate the percent of indorsement by each criterion sub-group and the difference between these percentages for each factor – the item discrimination indexes.

3. From the data of Group 3 in Study 2, compute the percentage indorsement indexes for the first alternative of each item under straight-take and fake-take conditions.

4. Choose response categories for inclusion on one or another of the five "Preliminary Empirical Keys" if:

   (a) the discrimination index was larger for that factor than for any of the others,

   (b) this largest discrimination index was significant at or beyond the .05 level (with $N_1+N_2 - 2 = 80$ df, a discrimination index of 10% at an indorsement index $\geq$ 95% or $\leq$ 5% and a discrimination index of 22% for a 50% indorsement index are significant at this level), and

   (c) the indorsement indexes for the two conditions were approximately equal or, if they differed, that for the fake-take was closer to 50% than that for the straight-take.

Although the procedure described was used as the primary basis for item keying, other data were available and were utilized in some instances to confirm decisions made or to decide marginal cases. These data included discrimination indexes computed from the ROTC sample data (except those for factor IV) and indorsement indexes based on the Fraternity sample, the ROTC sample, and Groups 1 and 2 from the Student Officer and Air Cadet OCS-desirability study (Study 2).

15

Whereas the scores from the *a priori* keys were systematically ipsatized by counterbalancing stems from the several factors within paired items, the empirical keys are considerably less so. About 55% of the response categories scored on the empirical keys involve no scoring of the other alternative on any other key. Although the *presentation* of the items is still in a forced-choice format, the *keying* is thus considerably less "forced." Table 6 summarizes the essential itemetric characteristics of these preliminary empirical keys for the DAI.

TABLE 6. Itemetric Characteristics of the Preliminary
Empirical Keys for the DAI

Key for Factor	Nr of Keyed Responses	% Discrimination Indexes[*]		% Indorsement Indexes[*]		% Nonipsative Keyed Responses
		Median	Range	Median	Range	
I	33	19	14-30	52	21-83	58
II	35	17	10-27	56	26-89	71
III	35	20	14-36	50	21-84	34
IV	31	19	10-37	57	21-87	52
V	40	22	15-41	50	24-78	60

[*]Based on fraternity study sample (total N = 82).

No adequate cross-validation sample is yet available on which to estimate the concurrent validities of these keys. The preliminary sample of fraternity men used in developing these keys was clearly too small to partition it for this purpose (if, in fact, an N = 82 is large enough to justify its use at all for empirical key construction). It is for these reasons that these keys are labeled "Preliminary." It may be instructive, however, to indicate the degree of "recursive" validity for these keys when re-applied to the sample used in their construction. These values and the correlations among the set of keys and between these keys and the other rating variables based on the fraternity study sample are presented in Table 7.

TABLE 7. Correlations Among the DAI Preliminary Empirical Keys
and the Criterion Rating Factor Scores

(Estimated recursively[*] on the Fraternity Validation sample, N = 82)

DAI Prelim. Empir. Keys	Factor Scores					DAI Preliminary Empirical Keys				
	I	II	III	IV	V	I	II	III	IV	V
I	(.69)	.06	-.25	-.08	.16					
II	-.09	(.50)	.11	.15	-.22	-.26				
III	-.34	-.14	(.63)	-.20	-.14	-.40	.07			
IV	.15	.21	-.31	(.50)	-.01	.07	.03	-.55		
V	.15	-.01	-.23	.06	(.64)	.21	-.40	-.56	.21	

[*]Not independent, cross-validation estimates since based on the same data used, in part, to build these keys.

16

The crucial question, of course, is how much shrinkage in these empirical key validity estimates (the values inclosed in parentheses in Table 7) can be expected on cross-validation. It seems unlikely that the ultimate validities will drop as low as those for the original *a priori* keys presented in Table 5.

The other feature of Table 7 worth noting is that 4 of the 10 interscale correlations (I-III, II-V, III-IV, and III-V) are moderately high negative. This is due (at least in part) to the fact that on a number of items both response categories are keyed, but for different factors. These four pairs of keys have respectively 7, 7, 10, and 5 such jointly keyed items whereas each of the · remaining six pairs of keys involve not more than three such joint keyings. To obtain mutual independence among any subsequent keys, it will be wise apparently to permit even fewer joint keyings than was done here. To more adequately map the rating factor correlations, it might even be well to key the *same* alternative on several items for both factor II and factor IV where the discrimination indexes warrant it.

THE OCCUPATIONAL PREFERENCE INVENTORY (OPI)

The development of this instrument parallels very closely that of the DAI, both in terms of rationale and in the kinds of data collected and used in constructing the test and its scoring keys. The major departures consist in the class of stimuli used and in the length of the test that was built.

From the occupational titles listed in the Holland Vocational Preference Inventory, a subset of 164 was selected on a judgmental basis as reflecting personality attributes like those indicated by the five rating factors. The question asked when considering a given occupation was "Would an expressed preference for this occupation be particularly likely for a person who was high or low on one of the factors?" This proved to be an extremely difficult judgmental task and not much confidence was placed in the factor designations arrived at for the occupational titles selected for further study.

As with the adjectives, single-stimulus OCS-desirability ratings were obtained on the 164 occupations from the sample of college students (Study 1). Contrary to the high correlation between the two sex groups' mean desirability values for the adjectives (.98), that for the occupations was only .87. Consequently, only the males' ratings were again used to pair the occupations into forced-choice sets using the same criteria employed with the adjectives. A few additional occupations were added to those available from the Holland Inventory to balance out the factors poorly represented. The preliminary forced-choice form of the *Occupational Preference Inventory* (OPI) thus constructed consists of 60 binary, forced-choice items.

This instrument was included in the batteries for the Student Officer and Air Cadet OCS-desirability sample and the ROTC and Fraternity validation samples (Studies 2, 5, and 6). These data have been used to determine the concurrent validities of the *a priori* keys (those derived from the original staff judgments of factor relevance employed in constructing the forced-choice items) and as a basis for developing preliminary empirical keys for the five factors. The criteria and procedures used to construct these keys were the same as those described for the DAI. The concurrent validities and interscale correlations among the *a priori* keys of the OPI, based on both the ROTC and Fraternity Study samples, are presented in Table 8.

The concurrent validities of these *a priori* keys against the peer-rating factor scores are, with the exception of the key for factor V, essentially zero. These validities, however, should be interpreted in the light of the generally negative interscale correlations among those *a priori* keys relative to the mutually independent rating variables. If revised keys can be built for this instrument which effectively map the rating factor score correlations and which also utilize empirical item validity data, some improvement in concurrent validities should be expected. The amount of improvement, however, may not be very great, owing to the relatively small pool of items available from which to select response categories for empirical keying.

17

TABLE 8. Correlations Between the OPI A *Priori* Keys and
the Criterion Rating Factor Scores

OPI A *Priori* Keys	Factor Scores						OPI A *Priori* Keys				
	I	II	III	IV	V		I	II	III	IV	V
ROTC Sample (N = 84)											
I	(.09)	-.07	-.30	.10	-.19						
II	-.04	(.07)	.09	-.01	.13		-.49				
III	-.06	.00	(.05)	-.20	-.02		.08	-.43			
IV	-.00	-.20	.05	(.03)	-.15		-.26	-.09	-.30		
V	.04	.20	.12	.11	(.24)		-.16	-.09	-.46	-.21	
Fraternity Sample (N = 82)											
I	(-.07)	-.17	.18	-.10	-.17						
II	.20	(.17)	.04	.09	.28		-.44				
III	-.13	-.16	(.13)	-.01	-.30		.21	-.53			
IV	-.08	-.08	-.15	(.05)	-.10		-.35	-.23	-.17		
V	.04	.21	-.09	.15	(.25)		-.17	.02	-.56	-.22	

Note.— Numbers in parentheses are validity coefficients; correlations of .21 significant at the
.05 level, of .28 at the .01 level.

Empirical keys have been built for this inventory following the same procedures described
for the DAI. The degree of ipsatized scoring has been reduced by approximately the same amount
as with the DAI preliminary keys — about 57% of the response categories scored on these prelim-
inary empirical keys for the OPI involve no scoring of the paired alternative on another key.
Itemetric characteristics of these keys are presented in Table 9.

TABLE 9. Itemetric Characteristics of the Preliminary
Empirical Keys for the OPI

Key for Factor	Nr of Keyed Responses	% Discrimination Indexes[*]		% Indorsement Indexes[*]		% Nonipsative Keyed Responses
		Median	Range	Median	Range	
I	12	21	15-29	50	29-72	50
II	12	18	13-35	51	30-67	80
III	11	15	12-26	51	24-68	36
IV	10	17	13-24	51	30-77	40
V	15	25	15-36	55	38-72	67

[*]Based on Fraternity study sample (total N = 82).

18

As with the DAI, these keys were used to score the answer sheets from the fraternity sample subjects whose responses had been used as the primary basis in constructing these keys. The recursive (non-cross-validated) correlations among these keys and between these keys and the rating-factor scores are presented in Table 10.

TABLE 10. Correlations Among the OPI Preliminary Empirical Keys and the Criterion Rating Factor Scores

(Esti ated recursively* on the Fraternity Validation sample, N = 82)

OPI Prelim. Empir. Keys	Factor Scores					OPI Preliminary Empirical Keys				
	I	II	III	IV	V	I	II	III	IV	V
I	(.35)	-.04	.08	-.29	.30					
II	.11	(.47)	-.09	₁14	.07	-.22				
III	-.21	-.07	(.33)	-.07	-.26	-.06	-.35			
IV	-.15	.21	-.17	(.40)	-.14	-.78	.19	.07		
V	.21	.08	-.09	-.01	(.56)	.21	.34	-.63	-.18	

*NOT independent, cross-validation estimates since based on the same data used, in part, to build these keys.

While these recursive estimates of the concurrent validities for these keys are appreciably larger than for the corresponding *a priori* keys, they are not as large as those similarly obtained for the DAI preliminary empirical keys. Neither is it likely that use of these keys in conjunction with those from the DAI will much improve the predictability of the criterion ratings since the correlations between corresponding keys for the two tests approximate the recursive validity estimates of the OPI scales (i.e., .33, .23, .37, .25, and .63 respectively).

These facts, when considered together with the sizable shrinkage in the estimated concurrent validities expected on cross-validation, indicate a rather poor prospect for this test as a contributor to the prediction of the peer-rating criteria. Nonetheless, the OPI will be included in the major validation study battery to permit a more precise evaluation of its worth. If it seems to contribute independently, however little, to the reduction of estimation error, it may warrant further work devoted to lengthening the test and improving the format in order to capitalize more fully on the unique properties of these kinds of stimuli and response processes.

THE FORCED-CHOICE SELF-REPORT INVENTORY (FCSRI)

The general rationale and method which underlie the development of a forced-choice inventory of self-report statements are essentially the same as that employed in the construction of the two previous tests. The scope of the effort, however, and some of the procedural details are different.

The initial step was to compile a very large pool of self-report statements, from which to select test stimuli. Some 7000 such statements were drawn from previously published inventories and questionnaires. Sources from which items were drawn (including revised editions of some of these instruments) are the following:

1. Minnesota Multiphasic Personality Inventory
2. California Psychological Inventory
3. The Opinion, Attitude, and Interest Survey

19

4. Edwards Personal Preference Schedule
5. Guilford-Zimmerman Temperament Survey
6. Bell Adjustment Inventory (Student Form)
7. Bernreuter Personality Inventory
8. 16 Personality Factor Questionnaire (Forms A, B, and C)
9. Inventory of Factors GAMIN (abridged edition)
10. Inventory of Factors STDCR
11. Guilford-Martin Personnel Inventory
12. Maslow S-I Inventory
13. A-S Reaction Study
14. Study of Values
15. Hilden Universe of Personal Concepts
16. Woodworth Personal Data Sheet
17. California Test of Personality (Adult Form A)
18. Thurstone Personality Schedule
19. Minnesota Personality Scale (Men's Form)
20. Laird Personal Inventory
21. Minnesota T-S-E Inventory
22. Loofbourow-Keys Personal Index-Test 3
23. Lentz C-R Opinionnaire (Forms J and K)
24. Cornell C. S. I. (Form N)
25. Cornell Medical Index (Men's Form)
26. Coop. Inventory (H-Alb)

Many of the statements in this Master Item Pool (MIP) were judged to be irrelevant for our purposes (e.g., those dealing with purely medical history, those reflecting bizarre mentation) and were eliminated at this point from further consideration. A second criterion employed to reduce the MIP to a workable size was to eliminate items which were not stated in a self-referring grammatical form. If neither the subject nor the object of the statement was at least implicitly in the first person, the item was culled from the pool. The staff was cognizant of the possibility that a number of potentially useful items might be precluded from further study by this process, e.g., those apparently asserting or denying matters of fact about which persons might differ in their beliefs or opinions. But in view of the over-abundance of more directly relevant stimuli still contained in the pool, these items were set aside for the present for possible later use in another form of test.

The final step in reducing the item pool to a manageable size was the result of our attempt to categorize the remaining items in terms of their apparent relations to the five rating factors. Three judges, each thoroughly familiar with the five factors and the set of rating scales known to load highly on them, independently sorted each of the items remaining in the pool into one of 10 categories (5 factors x 2 poles) in terms of their evaluations of the content of these items. The task for the judges was to decide which pole of which factor would be indicated by a self-indorsement to each item. Those items which the judges felt they could not categorize and those about which there were strong differences of opinion which could not be resolved in joint discussions were also dropped from further study.

At this point there remained some 1606 self-report statements in the pool, each of which had been sorted into one of the 10 factor-pole categories by the judges. While all exact duplicates had been culled from the pool as it was being compiled, there remained a fairly large number of "functionally synonomous" statements — i.e., items which, while they differed slightly in grammatical form or in the exact words used, seemed to assert the same thing.

To facilitate the construction of matched-content equivalent forms for subsequent data collection and as a further check on the factor-pole sorting, each of the ten sets of items were "detail sorted" into more homogeneous subcategories. The systems of subcategories were evolved in the process of sorting in that whenever an item was encountered which seemed to reflect a slightly different nuance or aspect of the factor-pole than was implied by any of the subcategories already established, it was made the nucleus of a new one. The critical descriptive terms in the statements in each subcategory were then listed and any additional synonyms which could be thought of (or in some cases, any which could be found in the Thesaurus) were added in an attempt to define the focus and the scope of each subcategory.

The items in each subcategory were then separated into four piles as evenly as possible. Each of the piles in each subcategory was assigned to one of four forms of the *Self-Report Item Pool* (SRIP, Forms A, B, C, and D). In preparing these forms (which comprise 402, 400, 403, and 401 single statement items, respectively), each deck of item cards was thoroughly shuffled and the items were then numbered sequentially and typed in booklet form on mimeograph stencils. Separate cover sheets for use under standard self-report instructions and for use in collecting OCS-desirability ratings were prepared.

Although these single-stimulus inventories are not intended for use as selection instruments, it may be interesting to examine some properties of the responses to them anyhow. Since each item included in each form has been judged to be relevant to one or another of the factors, *a priori* scoring keys could be constructed for each factor on each test form. Self-report data (true-false responses) were obtained from 55 of the original 82 men in the fraternity validity study. These 55 men are highly representative of the original total group in the sense that the form of the criterion rating score distributions, the medians, and the ranges on all five factors for these 55 men are very similar to those based on the total group of 82. These respondents were scored on all 20 of these keys. The equivalent-form reliabilities for these scales and the correlations of these scales with the criterion ratings were computed and are presented in Table 11.

The relatively high equivalent-form reliabilities indicate that the 1606 items were fairly well distributed over the four forms — at least for factors I, II, III, and IV. The somewhat lower values for factor V keys may well be a function of the relatively small number of items in these keys. The validities for these *a priori* keys, while not high, are consistently above zero and of about the same magnitude as those for the *a priori* keys of the DAI.

Although the size of the subsample from the Fraternity Study from whom SRIP answer sheets were obtained was small (N = 55), some use was made of the criterion factor data for these subjects in constructing preliminary empirical keys for these tests (and, as will be discussed later, in pairing statements for use in a forced-choice inventory). With these few cases, of course, estimates of itemetric characteristics (indorsement indexes and, even more so, discrimination indexes) are quite unstable, and no great faith was put on the replicability of these exact values nor on our ability to cross validate keys based upon them. Rather, these indexes were computed and preliminary empirical keys were built primarily to assuage our curiosity about some other questions.

First we were curious to see how well our subjective judgments of factor relevance for the items compared with indexes based on the actual performance of criterion cases (however scant the data may have been). Secondly, we wanted to see if keys based on test stimuli such as these could be made to yield recursive estimates of concurrent validity as high (or higher) than those obtained using adjectives (Table 7) or whether they would be somewhat lower — for instance, more like those obtained with the occupational titles (Table 10).

TABLE 11. Validities and Equivalent Form Reliabilities for the SRIP *A Priori* Keys

Factor	Form	Nr of Keyed Responses			Validity	Equivalent Form Reliabilities Form			
		T	F	Total		A	B	C	D
I	A	57	45	102	.27				
	B	57	45	102	.30	.89			
	C	58	44	102	.22	.79	.85		
	D	58	45	103	.24	.80	.85	.87	
II	A	32	48	80	.45				
	B	31	48	79	.40	.82			
	C	32	48	80	.34	.74	.79		
	D	31	47	78	.27	.73	.78	.88	
III	A	45	43	88	.25				
	B	44	43	87	.29	.88			
	C	45	44	89	.24	.80	.83		
	D	44	44	88	.33	.85	.85	.85	
IV	A	35	58	93	.41				
	B	36	57	93	.23	.88			
	C	36	57	93	.18	.84	.89		
	D	36	57	93	.22	.84	.86	.89	
V	A	26	13	39	.25				
	B	26	13	39	.34	.75			
	C	26	13	39	.41	.56	.56		
	D	26	13	39	.37	.65	.64	.70	

Note. — Validities and reliabilities estimated from 55 men in the Fraternity validity study.

Preliminary empirical keys for the five factors were built for only Forms A and B of the SRIP. The itemetric characteristics of these keys are presented in Table 12.

The correlations among the keys on each form, those between the two sets of keys, and those between each set of keys and the criterion rating factor scores are given in Table 13.

There are several features of Table 13 that warrant special mention. The first is that the recursive validity estimates (in parentheses) are of the same order of magnitude as those obtained for the preliminary empirical keys of the DAI. Thus self-report statements and personality-descriptive adjectives appear to possess about the same potential for tapping the factors — each being somewhat more effective than occupational preference stimuli as used in the OPI. The second point is that the equivalent form reliabilities for these keys (in brackets) are only very slightly lower on the average than those between the corresponding *a priori* keys (see Table 11) for these two forms — this despite a marked reduction in the number of scored responses on the preliminary empirical keys.

The final feature worth noting is the pattern of intercorrelations among the keys for each form. These inter-key relationships approximate very closely the pattern of correlations among the rating-factor scores reported in Table 3 for the total Fraternity Study sample. Except for the moderately positive values between the keys for factors II and IV, the rest are essentially independent of each other. These keys involve no items which are jointly keyed (either in the same

TABLE 12. Itemetric Characteristics of Preliminary Empirical Keys for SRIP Forms A and B

	Key for Factor	Nr of Keyed Responses	% Discrimination Indexes [*]		% Indorsement Indexes [*]	
			Median	Range	Median	Range
SRIP-A	I	26	21	19-31	45	11-82
Prelim.	II	35	24	20-39	44	11-80
Empir.	III	38	24	19-39	53	9-87
Keys	IV	37	24	21-36	53	11-89
	V	34	24	20-41	49	11-86
SRIP-B	I	19	24	20-30	55	18-82
Prelim.	II	33	24	20-42	38	9-78
Empir.	III	24	24	20-34	55	25-85
Keys	IV	28	24	20-38	56	22-84
	V	25	26	20-49	44	11-84

[*] Based on Fraternity Validity Study subsample (total N = 55)

or in the opposite direction) on two or more scales and hence the correlations involve no built-in scoring artifacts of these sorts. It thus appears that stimuli of this kind, when presented singly with a true-false response format, can be keyed so as to map the criterion dimensions rather closely. Whether similar results can be obtained when a forced choice format is used remains to be seen.

The next step was to construct a forced choice inventory using these self-report statements. Two forms of the *Forced Choice Self-Report Inventory* (FCSRI) have been built — Form A from the stimuli contained in SRIP-A and Form B from those in SRIP-B. The data used to pair statements were those obtained from the 55 subjects from the Fraternity Study subsample (Study 6) and the "admission-to-OCS-desirability" ratings of the statements in SRIP-A and SRIP-B obtained from 58 of the student officers and air cadets in Study 3.

From the Fraternity group, discrimination and indorsement indexes for each item (used above to construct the preliminary empirical keys for SRIP-A and B) were calculated. From the Study 3 ratings, the mean and standard deviation for each item's distribution were computed. In addition, the correlations between rating distributions for blocks of 60 (terminal blocks of 40) items within each SRIP form were computed. This was done to permit matching of items not only on centrality and variability parameters but also in terms of a high degree of correspondence between *individual* judges' ratings of the items.

The method used to pair the statements was as follows:

1. Consider a block of 60 (or 42 or 40) items from one of the SRIP forms.
2. Sort these items into 10 categories (5 factors x 2 poles) in terms of the original content judgments of factor relevance.
3. Pair items so that for each pair: .
    a. mean desirability values are nearly equal.
    b. standard deviations of the desirability judgments are nearly equal.
    c. correlations of desirability judgments are "high".
    d. indorsement indexes are nearly equal.
    e. factor relevance is different, i.e., [1] content judgments of the two items place them on similar poles (+ + or - -) of two different factors, and [2] empirical discrimination indexes are not high and in the same direction on the same factor.

TABLE 13. Correlations Among the SRIP Preliminary Empirical Keys and the Criterion Rating Factor Scores

(Estimated recursively* on the Fraternity Validation subsample, N = 55)

Factor	Factor Scores					SRIP-A Preliminary Empirical Keys					SRIP-B Preliminary Empirical Keys				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
SRIP-A Prelim. Empir. Keys															
I	(.71)	.21	-.30	-.05	.27										
II	-.03	(.64)	.05	.32	-.18	-.04									
III	-.27	-.05	(.58)	-.01	.00	-.31	-.13								
IV	.02	.31	-.09	(.63)	-.07	-.05	.48	-.28							
V	.16	-.00	.17	.01	(.73)	.16	-.03	.02	.15						
SRIP-B Prelim. Empir. Keys															
I	(.72)	.05	-.35	.01	.29	[.78]	-.19	-.15	-.01	.31					
II	.01	(.61)	.03	.30	-.09	-.02	[.87]	-.17	.46	-.01	-.22				
III	-.24	-.01	(.60)	-.15	.01	-.21	.05	[.75]	-.27	.07	-.18	.05			
IV	-.00	.27	.02	(.59)	.02	-.13	.43	-.16	[.82]	.19	-.05	.49	-.26		
V	.08	-.06	.05	-.12	(.63)	.17	-.14	-.07	.00	[.67]	.20	-.09	-.08	-.04	

*NOT independent cross-validation estimates since based on the same data used to build these keys; entries in parentheses ( ) are validities, those in brackets [ ] reliabilities.

24

4. When the criteria of step 3 can no longer *all* be met by any additional pairs of stems within a block, stop, and place the remaining stems in a set of 10 residual factor-pole categories.

5. Repeat steps 2 through 4 for all blocks of items in SRIP Forms A and B.

6. For the items from each SRIP form in the residual categories, on which correlations between rating distributions are either low or unavailable (if the items are from different blocks), form additional pairs according to criteria 3(a), (b), (d), and (e).

7. For all remaining items, relax the criteria in step 3 as required to form additional pairs.

8. Order the stems within pairs and the paired items within the forced choice form so as to break position and content response sets. Intermingle plus-plus and minus-minus desirability items throughout each forced-choice form.

The rationale for the criteria in step 3 above should be fairly obvious. Insofar as one can depend at all on single-stimulus item parameters in the construction of tightly matched forced-choice items, these criteria should be sufficient. If the effects of changing contextual format and the respondents' task markedly change the stimulus properties of the individual stems, then these procedures will not be sufficient for the purpose. In any event, a second check on the fakability and empirical validities of the *paired* items is certainly required before final scoring keys are formulated for these instruments.

Perhaps someday we will have learned enough about the effects of going from single-stimulus formats to those of forced-choice presentations to be able to specify more adequately which kinds of stems we should pair with which others, but that day has not yet arrived. All we can do at present is to pair items as tightly as possible in terms of the kinds of criteria listed above, hope for the best, and discard those forced choice items that fail to possess the desired properties. It is because of the anticipated need to disregard sizable numbers of items in the construction of scoring keys that so many binary items were constructed and placed in FCSRI forms. Form A contains 192 binary forced-choice items and Form B, 199. If no joint keying of any item is done and if as many as 60 or 70 items in each form are unusable on any key, one should still have about 30 scored responses for each factor key for each test form. Previous results indicate that keys of this length are adequate to achieve fairly high reliabilities, but, if desired, one could always combine the data from the two forms to obtain more precise measures.

A set of preliminary keys have been constructed for the two forms of the FCSRI, based primarily on the data from the 55 men from the Fraternity Study subsample who took the SRIP forms. These data, of course, come from single-stimulus presentations and these preliminary keys will in all probability require extensive revision once empirical itemetric data based on the FCSRI forms themselves become available. These keys will be used to determine the degree to which single stimulus data can be used to construct effective scoring keys for forced-choice inventories of this sort. Table 14 indicates the number of scored responses on each of these keys and the percentage of nonipsatively keyed responses on each of them.

On the average, these keys are slightly less highly ipsatized than the preliminary empirical keys for the DAI and OPI reported above. On Form A only the I-III and I-V key combinations involve more than two jointly keyed items (6 and 4, respectively); and on Form B, the I-III, II-V III-IV, and III-V combinations exceed the arbitrary minimum of two joint keyings (4, 4, 7, and 4 respectively).

TESTS SELECTED FROM OTHER SOURCES

THE WELSH FIGURE PREFERENCE TEST (WFPT)

We will now turn our attention to a test which, while it was originally built for other purposes, has been used in some of our studies.

TABLE 14. Number of Keyed Responses and Percentage of Nonipsatively
Keyed Responses on the Preliminary FCSRI Factor Keys

Key for Factor	FCSRI-A		FCSRI-B	
	Nr of Keyed Responses	% Nonipsative Keyed Responses	Nr of Keyed Responses	% Nonipsative Keyed Responses
I	25	52	24	63
II	24	75	29	72
III	28	57	30	43
IV	32	81	21	62
V	27	63	33	70

The research edition of this test, developed by George S. Welsh, consists of a series of 400 designs printed in booklet form with eight designs per page. For each of the designs a subject indicates on a separate answer sheet whether he likes or dislikes that particular figure.

On the basis of arguments and some preliminary data presented by Welsh (1959), a set of eight scoring keys were selected as possibly related to one or another of the five rating factors. These included the three "validating" scores — Don't Like (DL), Repeat (RP), and Conformance (CF); four of the "empirically derived" scales — Barron-Welsh Art Scale (BW), Revised Art Scale (RA), Male-Female Scale (MF), and Neurophychiatric Scale (NP); and one of the "judged" item scales — Movement (MV).

The answer sheets from both the ROTC and the Fraternity validity studies were scored on these eight keys and the results were correlated with the criterion rating factor scores. In neither study did any of the 40 correlations between these keys and the criterion ratings exceed .19 in magnitude, nor was there any evidence of a match between our conjectures of factor relevance for these keys and the obtained values.

Since the stimuli of this test are not obviously relatable to the content of the five factors, and since the pool of stimuli is large (400 designs), it was felt that useful empirical keys might be developed for this instrument. Thus an item analysis of the test responses based on the fraternity sample was undertaken. Percentage discrimination indexes on each factor were computed for each item. Keys were formed for each factor by choosing those items whose discrimination indexes for the factor were larger than for any other factor and greater than some arbitrary minimum. The only other restraints were (1) that no item be keyed on more than one factor, and (2) that a roughly equal number of responses be included in each key. Two sets of keys were developed — one corresponding to a minimum discrimination index of 15% and one to a minimum value of 20%. Itemetric characteristics of these keys are given in Table 15.

It can be seen from Table 15 that empirical keys of reasonable length with acceptable item characteristics can be built by this procedure. The most striking feature of these keys, however, is the variation in the percentage of "Like" responses. Keys for factors I and II contain practically no items keyed in the "Like" direction, while keys for factors III and V are composed predominantly of such items. Only the keys for factor IV are reasonably balanced with respect to "Like" (acquiescence?) responding. Since the overall "Dislike" score (DL) showed no appreciable validity on any of the five factors (a coefficient of .15 for factor I in the Fraternity Study was the highest), it seems that liking or disliking of *some specific characteristics of only some of the figures* is what is relevant in distinguishing highs and lows on the various factors.

TABLE 15. Itemetric Characteristics of WFPT Preliminary Empirical Keys

Key for Factor	Nr of Keyed Items	% Discrimination Indexes[*]		% Indorsement Indexes[*]		% "Like" Responses Keyed
		Median	Range	Median	Range	
20% Keys						
I	27	23	20-37	54	29-77	0
II	18	22	20-32	42	19-81	6
III	22	25	20-30	48	24-82	77
IV	23	22	20-30	46	27-76	39
V	23	22	20-32	46	24-77	91
15% Keys						
I	52	20	15-37	47	21-77	2
II	35	20	15-32	46	19-80	6
III	50	19	15-30	44	14-82	74
IV	53	19	15-30	48	17-80	51
V	42	20	15-32	46	17-77	88

[*] Based on Fraternity Study sample (total N = 82).

Before a detailed analysis of the stimulus properties of the figures scored on particular keys is undertaken, however, it has been deemed advisable to wait for the availability of cross-validation data.

The subjects in the Fraternity Validation Study were scored on both sets of these preliminary empirical keys and the relations among these keys and with the rating factor scores, including recursive validity estimates, are presented in Table 16.

Several features of the data presented in Table 16 deserve special comment. First the recursive validity estimates (in parentheses) are of about the same magnitude as were obtained with the OPI for similarly constructed keys but lower than those for the preliminary empirical keys for the DAI and SRIP forms. While the smaller values for the OPI may have resulted from the relatively small number of response categories available from which to select keyed responses, this is not a tenable explanation for the WFPT where the number of response categories available is as large or larger than for the DAI and SRIP forms. It would seem instead that the greater adequacy of the keys constructed for the latter tests can be attributed to the greater content relevance of the stimuli used in these instruments.

The second point to be made refers to the pattern of interkey correlations — especially those between the keys for factors I and II — in each set. Bearing in mind that the rating scores for these two factors (used as a basis for designating members of the contrasted criterion groups for selecting items for these keys) are essentially uncorrelated, these high interkey correlations are rather surprising. Referring to Table 15, it can be seen that these keys are the ones which are based primarily on "Don't Like" responses. Therefore the magnitude of these correlations is, in all probability, primarily due to some common response set operating over the set of stimuli on which these keys are based. The moderately high correlations between the keys for factors III and V is presumably due, at least in part, to the operation of the opposite kind of set.

The final comment relative to Table 16 is by way of a caution. The very high values between corresponding keys in the two sets should not be interpreted as any kind of reliability

27

TABLE 16. Correlations Among the WFPT Preliminary Empirical Keys and the Criterion Rating Scores
(Estimated recursively* on the Fraternity Validation Study sample, N = 82)

Factor	Factor Scores					WFPT 20% Preliminary Empirical Keys					WFPT 15% Preliminary Empirical Keys				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
WFPT 20% Prelim. Empir. Keys															
I	(.44)	.06	-.16	-.03	.10	.74									
II	.26	(.31)	-.06	.11	-.02	-.30	-.11								
III	-.18	.16	(.54)	-.02	.16	-.12	-.18	-.13							
IV	-.00	.24	-.06	(.42)	-.05	-.42	-.32	.43	-.28						
V	.03	.07	.09	-.10	(.51)										
WFPT 15% Prelim. Empir. Keys															
I	(.40)	.04	-.12	-.03	.07	.97	.80	-.24	-.16	-.39	.82				
II	.24	(.22)	-.06	.03	.01	.76	.97	-.12	-.24	-.31	-.30	-.16			
III	-.21	.15	(.52)	-.01	.15	-.37	-.16	.95	-.14	.49	-.32	-.40	-.14		
IV	-.07	.20	-.03	(.44)	-.05	-.27	-.33	-.14	.94	-.21	-.38	-.33	.50	-.17	
V	.02	.07	.08	-.01	(.51)	-.41	-.33	.44	-.24	.96					

*NOT independent, cross-validation estimates since based on the same data used to construct these keys.

estimates since there is a sizable amount of item overlap between these keys. Every item in a given 20% key is also scored on the 15% key for that factor.

In summary, our evaluation of this test and its potential usefulness for our purposes is that only if empirical keys based on more stable itemetric data can be built which cross-validate adequately and which possess relatively low correlations with other predictor scales for the factors, will it deserve further consideration. It does not appear at the present to be anywhere near as suitable as some of the other instruments based on more directly relevant stimulus materials.

CATTELL'S 18 FACTOR OBJECTIVE-ANALYTIC TEST BATTERY – GROUP FORM

Cattell has published a battery of some 44 short, so-called objective tests for the assessment of personality factors. A careful reading of the handbook, conversations with Cattell, and correspondence with him and others working with these instruments indicate that a poor match exists between Cattell's rating-scale factors and those arising from analyses of the Objective Analytic Battery. The individual tests, however, have never been given together with the rating scales in an attempt to determine if any subset of them relate to the rating-scale factors. In addition, Cattell does not publish normative data on the scales scored from these tests but leaves it to individual users to develop their own. Thus two preliminary studies were undertaken to provide some tentative information on the feasibility of using the tests in this battery.

In the first of these investigations (Study 4) the intent was to become familiar with the complex administrative and scoring procedures for these tests and to determine on which of them sufficient variability could be obtained to permit correlational analysis against the rating variables. On the basis of these data certain of the tests were eliminated from further consideration, and for others instructions and response formats were modified to simplify further work with them.

The second study in which portions of the O-A Battery were used was the Fraternity Validation Study. The 82 participants were administered 30 subtests which had been selected on the basis of group administrability, ease of scoring, judgments of relevance of the Master Index (MI) variables scored on each to the rating factors, and adequacy of the norms generated in the pretesting study. Scores on 58 MI variables were obtained on each man in this study. Of these 16 correlated significantly with one or more of the criterion rating variables. The highest single zero-order validity was .47 for MI 117 against the factor V ratings. The validities for each of the 58 MI variables against each of the five factors are given in Table 17.

In an additional analysis, all 58 of the MI variables were run as independent variables against each of the five criterion rating factors using a stepwise multiple regression program on the IBM 704. A significance level of .50 was set for adding or deleting a predictor and the program was run until no further changes were called for at this level. The results as of the terminal stage for each of these analyses are summarized in Table 18.

The multiple correlations presented in Table 18 are, of course, highly inflated due to the large amount of error-fitting permitted by the small sample and the extremely liberal significance level specified. The primary intent of these analyses was, however, not to estimate such indexes, but rather to provide some basis for further reducing the number of tests from the O-A Battery to be included in subsequent studies. The level of significance used was chosen so as not to exclude from further consideration any test which might have some predictive potential. On the basis of these results and the values presented in Table 17, the number of subtests from the O-A Battery to be included in the major validation study was reduced to 23, yielding scores on 47 of the MI variables, and resulting in a reduction of about 25% in test administration time.

THE ESOTERIC KNOWLEDGE TESTS

The rating scales which load highly on Factor V suggest a picture of the highly-rated individual as one who is interested and knowledgeable in arts and letters, who is sophisticated

TABLE 17. Product-Moment Correlations Between 58 MI Variables and the Criterion Rating Factors

MI	Rating Factor				
	I	II	III	IV	V
150b	-.14	.01	.11	-.07	-.11
7	.04	.16	-.24*	.06	.14
193	-.01	-.06	-.14	.07	-.11
67	-.01	-.12	-.04	.03	-.11
307	.16	-.01	-.22*	-.09	-.03
117	.13	-.04	-.15	-.09	.47***
309	.09	-.07	-.15	-.16	-.10
237	.06	.07	.14	.23*	.12
287	-.18	.12	-.03	.09	.21
134	.01	.09	-.05	.02	-.17
25	.12	.13	.02	-.10	.26*
53	.21	-.04	-.19	.09	-.10
314	.03	-.11	-.15	-.03	.20
147	.00	.11	-.03	.03	.04
288	.08	-.09	-.12	-.17	.06
308	.01	-.14	-.04	-.14	-.02
34	.18	-.11	-.15	-.05	.10
271	.08	-.02	-.05	.08	.00
108	.03	.06	-.04	-.06	.05
145	.11	-.05	.03	.00	.03
101	-.01	-.05	-.10	-.14	-.10
194	.04	.02	-.03	-.01	-.01
152	-.20	.01	-.07	.16	-.20
327	.02	.06	-.10	-.05	.07
316	-.06	.25*	.00	.17	-.09
191	-.12	-.10	.03	-.10	.09
123	-.07	-.12	.02	-.15	.02
97	-.14	-.01	.17	-.15	.14
146	.04	.19	-.08	-.03	-.05
211	-.10	-.16	-.01	-.06	-.14
283	.00	-.10	.02	-.12	.13
38	-.09	.00	-.20	.12	-.12
35	-.07	.02	-.14	-.06	.07
246	-.25*	-.16	.23*	-.06	-.31**
219	.20	-.08	-.06	-.06	.24*
6	.13	-.10	-.18	-.07	.07
9	-.02	.02	-.06	-.12	.23*
159	-.33**	-.08	-.12	.09	-.30**
151	.05	-.07	-.02	-.06	.10

Note.— Correlations based on Fraternity Study sample (N = 82).

* = Significant at .05 level

** = Significant at .01 level

*** = Significant at .001 level

(Table continues on next page)

TABLE 17 (Continued)

MI	Rating Factor				
	I	II	III	IV	V
112	.13	-.16	-.20	.07	.07
100	.16	.04	-.04	-.05	.11
103	.02	.02	-.18	.00	-.14
206	.01	.07	-.25*	.12	-.07
203	-.15	-.07	-.05	.07	-.02
31	.06	-.05	.11	-.04	.01
36	.19	-.04	-.07	-.09	.15
280	-.07	-.02	.23*	.00	.00
330	-.10	-.07	.04	-.13	.00
199	-.12	-.15	.00	-.03	-.16
275	.06	-.03	-.22*	.04	-.02
87	-.16	-.08	-.08	-.02	-.26
102	.21	.07	-.03	.04	.26*
307 and 308	.13	-.02	-.14	-.03	-.12
167	.11	-.09	-.21	-.06	-.39***
133	.04	-.09	-.07	.03	.07
109	.02	-.04	-.06	.08	.07
110	.05	.18	.11	-.08	.12
275	-.13	-.11	.05	-.12	.21

* = Significant at .05 level
** = Significant at .01 level
*** = Significant at .001 level

TABLE 18. Summary of the Terminal Multiple Regression Functions for the Five Factors as Predicted by the 58 MI Variables

(Fraternity Validity Study, N = 82, $a = .50$)

Analysis for Factor	Nr of Predictors Included	R*
I	28	.82
II	30	.75
III	28	.76
IV	37	.83
V	27	.87

* NOT cross-validation estimates.

and urbane, and who is curious and attracted by the unusual or exotic features of the environment. This suggested that one way to tap Factor V variance would be to assess the amount and the diversity of information possessed by the subjects concerning obscure or esoteric areas of knowledge. Such tests should be suitable for use in selection contexts since the assessee is instructed to do the best he can and if items are carefully constructed, he ought not to be able to fake such a test in a positive direction.

Two tests of this type were included in the battery administered to the subjects in the Fraternity validity study:

> *General Knowledge – A.* This test consists of 112 multiple choice items drawn from the Coop General Culture Test – Form X and spans all six areas of academic knowledge tapped by that test. A single score (total number of items correctly answered) was obtained.

> *Culture – E:* This test consists of 30 multiple choice items for which the subject must identify the esthetic or cultural area most closely associated with each of the 30 stimulus words.

A number of other tests included in the battery – some from the O-A set (e.g., MI 117 "Highbrow Tastes" from test G27) and some used primarily to assess other attributes such as risk taking tendencies – could also be scored for breadth or amount of knowledge or cultural sophistication possessed by the subject. These two tests, however, were judged to provide the clearest test of the rationale underlying this method for assessing factor V because of the high specificity and relative obscurity of the information needed to score well.

The validities against factor V rating scores were .47 for General Knowledge – A and .33 for Culture – E. Unfortunately, these two test variables and most of the others in the battery which correlate appreciably with factor V ratings also relate rather highly one with another. Thus the amount of improvement in predictability of the factor V criterion that can be obtained by multiscale methods may be somewhat more limited than for the other criterion dimensions. It may be possible by item analysis methods to improve the validity of the General Knowledge key, but Culture – E is probably too short to permit any refinement by dropping low discriminating items from the keyed set without a marked drop in reliability. Both of these tests and at least one device not used in any of our preliminary studies (Mednick's Remote Associates Test of creativity) will be included in the battery for the major validation study.

THE RISK-TAKING MEASURES

Certain aspects of factors I and III seem to deal with generalized dispositions toward or against taking risks or chances. The "adventurousness" or "boldness" of factor I and the "impulsiveness" and "rashness" of factor III (minus) both seem to connote such tendencies, though perhaps with slightly differing manners of expression. A variety of maximum performance tasks can be constructed which would appear to elicit responses dependent on such dispositions. Gambling or betting situations, penality-for-guessing scoring systems for a variety of ability tests, or prediction-of-success measures on achievement tasks are a few that come easily to mind. Several tests and scoring methods of this kind were included in the Fraternity validity study battery.

> *Bet Preference Test.* This test is composed of 50 sets of four 2-choice bets. The subject is asked to rank the bets in each set in terms of his preferences for playing them. All bets have zero expected value but are varied within a set on either probability of winning or on the variance of the payoff. Scores are derived for both probability preferences and for variance preferences.

32

Self-Crediting Test — V. The subject is shown a relatively easy vocabulary question and told that he will be given a test containing more items of similar difficulty. He is then asked to set the amount of points (between 1 and 10) which he wants each question on this test to be worth. If he answers a question correctly he will receive the amount of points he was willing to risk for each question. If he does not mark the correct answer he loses as many points as he has made each question worth. After the first 10-item test, he is told that the next test will be a more difficult one and he is to reset the number of points he wishes to give these harder questions. A total of four tests are given; before each, the subject is asked to reset the value for the items on the next, harder, test. Holding knowledge constant, it is hypothesized that the high risker will want the next test to be worth more points than the low risker does.

Word Meanings. The subject is asked for each of 10 groups of 15 terms which ones belong in a stated category (such as musical terms). The terms are quite ambiguous and one would predict that high riskers would more readily include an ambiguous term. The test is repeated a second time with the introduction of a "penalty" for wrong answers and one can observe the number of terms originally included in the category under no-penalty conditions which the subject does not want scored under the penalty condition.

Verbal Intelligence (Test Risk). This test consists of 43 multiple choice vocabulary items. The subject is told that he will receive one point for every wrong alternative which he correctly identifies as wrong, but that he will lose three points for marking the right answer as being wrong. He is told that he may mark as many alternatives wrong for each item as he wishes. After the test is taken, the subject is asked to look back at his answer sheet and pick the right answer for each item out of alternatives which he has previously left blank.

A person who has previously marked two alternatives out of four as being wrong, when now asked to pick the correct answer from among the remaining two has a 50% chance of being correct if he has absolutely no knowledge about these answers and ventures a complete guess. The extent that this person is correct better than 50% of the time in situations like these (or correct more than 25% of the time when choosing from among four possible alternatives, etc.) is an indication of how much knowledge he really has about these questions which he is not willing to risk using. It was predicted that low risk takers would have more knowledge on items they were not willing to guess at than would high risk takers.

Dot Estimation. The subject is given a short time in which to compare many pairs of squares. Each square has dots in it and S must mark which square in each pair has the most dots. It is hypothesized that low risk takers will tend to count the dots in each square to make sure they are right, while high riskers will act after only a quick glance at the figures, thus they will attempt to answer many more pairs in the time allowed.

Two other instruments, the USAF Life Experience Inventory and a peer-nomination rating scale whose poles were labeled "Loves to take risks. A daredevil" and "Cautious. Does not like to take chances. Avoids risky situations." were also included in the battery but were not considered as maximum performance or objective test measures of risk-taking propensity.

Some 45 variables, many of them systematically interdependent, were scored on the objective risk-taking tests. Table 20 lists some of the variables from this set and their validities against the five rating-factor criteria.

The validities of these scoring variables against factors I and III, while generally in the predicted directions, are not very large. This coupled with the fact that several of them have built-in dependencies (differences or ratios between more directly obtained scores) indicates little promise for these particular methods of assessing risk taking tendencies. In addition, several of of the scoring methods, especially on the Bet Preference Test, are quite involved and time consuming. Some of the more conventional scoring formulas for the ability-like tests (e.g., Number correct or measures reflecting confidence in ability to do well on such tests) do relate moderately to factor V ratings. As mentioned before, however, these scores also correlate quite highly with other valid measures of this factor (e.g., General Knowledge — A) and accordingly would not add much to the variance accounted for if combined with these other variables.

TABLE 19. Validities of Selected Risk Variables Against the Rating Factor Criteria

(Estimated from the Fraternity Validity Study Sample, N = 82)

Test	Scoring Variable	Validity				
		I	II	III	IV	V
Bet Preference Test	Av. Nr of times larger variance chosen	.12	-.05	-.13	-.03	-.07
	Av. Nr of times lower probability chosen	-.07	-.04	-.31	-.04	-.27
	Sum of riskier variance and probability prefs.	.05	-.08	-.34	-.06	-.27
Self Crediting Test – V	Total points earned	.19	.07	-.11	-.01	.46
	Total points risked	.17	-.00	.04	.11	.54
Word Meanings	Total Nr of inclusions (of 150 possible)	.11	-.05	-.05	-.03	.20
	Total Nr of inclusions reaffirmed under penalty condition	.04	-.05	-.01	.02	.11
	Difference between nr of inclusions and nr reaffirmed	.12	.01	-.06	-.09	.15
Verbal Intelligence (Test Risk)	Total Nr correct	.25	.04	-.21	.03	.49
	Total Nr misinformed (right answer called wrong)	-.16	-.10	.08	-.13·	-.21
	Standard of assurance or amount of information possessed but not used	.19	.02	-.29	.01	-.08
	Total Nr misinformed divided by total nr of alternatives called wrong	-.20	-.07	.09	-.06	-.44
Dot Estimation	Total Nr attempted	.07	-.05	-.20	.00	-.11

Because of the validities for Self-Crediting Test – V against factor V ratings in this sample, it will be included in the major validation study. A device recently developed called the *Decision Analysis Test* (DAT)[5] will also be included in the major validation study battery. It makes use of 4 by 4 matrix-game displays and asks the subject to rank order the rows in each matrix in terms of his preferences for playing each of them against a random opponent. The matrices are so constructed that there is an appropriate first choice on each of them for any subject wishing to play either a Laplacian, Optimist, Minimax, or Regret Strategy. Scoring keys for each of these strategies are available and the test yields four ipsatized scores (due to the ranking method of

---

[5] By William L. Hays and Robert L. Isaacson, University of Michigan.

data collection) for each subject. Some of the data from the Bet Preference Test indicated that such "pure strategy" variables (especially the Optimist and Minimax scores) might possess some validity for factors I and III.

GENERAL EVALUATION OF THE DEVELOPMENT STUDIES

Since so much of what has been presented in this report is based on small numbers of cases and since many of the results have not yet been cross-validated, there are few definitive conclusions which can be drawn. However, some tentative assertions do appear to be justified.

1. Five relatively independent and easily interpreted personality dimensions have been found to provide a basis for peers' perceptions of their close associates in numerous studies using samples from a variety of young adult male populations. The rating scales used to obtain these results are representative of a well-defined and extensive universe of personality attributes.

2. A variety of stimulus materials have been found or constructed which when properly presented to a subject, elicit self-report responses which can be scored to assess the subject's position on each of the five peer-rating dimensions with moderate accuracy. Formats for these tests have been constructed so as to minimize the effects of positive evaluative distortion tendencies. Various kinds of stimulus materials have been found to be differentially effective for the assessment of these factors.

3. Some maximum performance or "objective" test variables have been found which correlate with the criterion factors to a moderate degree. The most promising of these devices seem to be those based on the amount of esoteric knowledge possessed by the respondent when related to status on factor V. However, accurate predictions of the rating criteria, if attainable with such devices, will probably require multiscale methods.

THE FINAL BATTERY

The rating scales and tests we have developed or adapted will be administered to a new sample of between 400 and 500 college men composed about equally of students living in fraternity houses and in dormitories at the University of Michigan.

The following measures will be administered in group-testing sessions:

    20 Peer Nomination scales
    Cattell's Objective-Analytic Battery (22 tests)
    Culture — E
    Self-Crediting Test — V
    Remote Associates Test
    Decision Analysis Test
    General Knowledge — A

Four self-report inventories will be completed twice by each subject, administered as homework to be turned in at the next group-testing session. The first administration will be with standard self-report directions. The second will be with directions to answer as an applicant for the Air Force Officer Candidate School who wants to fake his way to acceptance.

    Descriptive Adjective Inventory
    Occupational Preference Inventory
    Welsh Figure Preference Test
    Forced-Choice Self-Report Inventory, Forms A and B

A double cross-validation analysis will be employed. Empirical scoring keys and multiple prediction functions will be constructed separately for each half sample. Validities and insensitivities to dissimulation will be estimated independently on the other half-sample.

REFERENCES

Cattell, R.B. Confirmation and clarification of primary personality factors. *Psychometrika*, 1947, 12, 197-220.

Cattell, R.B. *Personality and motivation structure and measurement.* Yonkers-on-Hudson, N.Y.: World, 1957.

Edwards, A.L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corporation, 1954.

Kaiser, H.F. Relating factors between studies based upon different individuals. Dittoed preliminary draft, Bureau of Educational Research, University of Illinois, July 1960.

Norman, W.T. *Problems of response contamination in personality assessment.* Lackland Air Force Base, Texas: Personnel Laboratory, Aeronautical Systems Division, May 1961. (Technical Note ASD-TN-61-43)

Tupes, E.C. *Relationships between behavior trait ratings by peers and later officer performance of USAF Officer Candidate School graduates.* Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center, October 1957. (Research Report AFPTRC-TN-57-125, AD-134 257)

Tupes, E.C. & Christal, R.C. *Stability of personality trait rating factors obtained under diverse conditions.* Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Center, May 1958. (Technical Note WADC-TN-58-61, AD-151 041)

Welsh, G.S. *Preliminary manual, Welsh Figure Preference Test, Research Edition.* Palo Alto, Calif.: Consulting Psychologists Press, Inc., 1959.

Wherry, R.J., Stander, N.E., & Hopkins, J.J. *Behavior trait ratings by peers and references.* Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Center, December 1959. (Technical Report WADC-TR-59-360)

INSTRUCTIONS TO RATERS

During the next hour you will be asked to describe some of the members of your group on a number of characteristics. Descriptions and ratings similar to these are common throughout industry, education, and military organizations. Almost all evaluations of other persons rely on such ratings. Therefore, the ability to judge others accurately is very important in many industrial, professional and military situations.

We want you to be as forthright and as accurate as you can in making your ratings. You may be assured that your evaluations will not be shown to any member of your group. Your ratings will be kept completely confidential.

Each of you has before you a roster of names of the persons in your group. In front of each name there is a number. On each page of the booklet you will be asked to rate the *other* members of your group on the characteristics described at the top of that page. Look at the sample page shown below:

```
                                                              (Scale #00)
    Group Roster #  _____    Rater #  _____

         A                                        B
    ┌─        ─┐                        ┌─          ─┐
    │          │                        │            │
    │          │                        │            │
    └─        ─┘                        └─          ─┘

               A    M    B

          1.   ___  ___  ___
          2.   ___  ___  ___
          3.   ___  ___  ___
          4.   ___  ___  ___
          5.   ___  ___  ___
          6.   ___  ___  ___
          7.   ___  ___  ___
          8.   ___  ___  ___
          9.   ___  ___  ___
         10.   ___  ___  ___
         11.   ___  ___  ___
         12.   ___  ___  ___
         13.   ___  ___  ___
         14.   ___  ___  ___
         15.   ___  ___  ___
         16.   ___  ___  ___
         17.   ___  ___  ___
         18.   ___  ___  ___
         19.   ___  ___  ___
         20.   ___  ___  ___
```

In place of the brackets you will find two contrasting characteristics described, one labeled "A" and one labeled "B". The numbers on the left of the page below the descriptions correspond to those on your group roster.

The procedures for making your ratings on *each page* are as follows;

1. In the spaces at the top labeled "Group Roster #" and "Rater #", write the roster number for your group and the number of your name on that roster. Do it now on the sample page above.

2. Find the line below the descriptions corresponding to the number of *your* name on the Group Roster and place an "X" on the middle space, (i.e. the space in the column under the letter "M".) Do it now on the sample page above.

3. Find the line numbered one larger than the number of persons in your group and draw a heavy horizontal line across the page through that number. (If there are twelve persons in your group *including yourself*, draw the line across each page through the number "13".) Do this now on the sample page by drawing a line across the page through the number ___. *Do not make any marks below this line on any page.*

4. Next, *READ CAREFULLY* the two descriptions labeled "A" and "B" at the top of the page. Choose the _____ persons in your group who are best described by the description labeled "A" and place an "X" in the space beside their numbers in the column labeled "A". Next, choose the____ persons in your group who are best described by the "B" description and place an "X" after their numbers under the letter "B". *DO NOT* rate any person in more than one column on any one page.

5. Rate all other persons in your group (whom you have not yet rated as either "A" or "B") under the column labeled "M".

6. Before you go on to the next page, check to be sure there is one and only one "X" after each number down to the heavy horizontal line. Also check to see that you have exactly _____"X's" in the "A" column and _____"X's" in the "B" column and that you have rated yourself as either "A" or "B".

In making your ratings use the special pencils provided. Place the "X's" on the short lines following the numbers under the letters "A", "M", or "B". *DO NOT* place them in the empty spaces between the columns.

Are there any questions?

Remember: Be as honest and accurate as you can. Make no omissions. Rate the specified number of persons as "A" and "B" on each page who come *closest* to being like the descriptions at the top. Do not be concerned about whether the descriptions fit these persons *exactly*, but only whether these are the persons *in your group* who are *most like* the "A" or "B" descriptions on that page.

Please work independently and quietly. Do not comment aloud but rather raise your hand if you have any questions as you proceed.

Turn the page now and begin.

Appendix (*Continued*)

RATING SCALES

Factor	Replicate	[A]	[B]
1	1	Talks a lot, to everybody.	Says very little; gives the impression of being occupied with thoughts.
I	2	Comes out readily with his real feelings on various questions; so that you know where you stand with him. Expresses his feelings, sad or gay, easily and constantly. Easy to understand.	Keeps his thoughts and feelings to himself. Often leaves you puzzled as to the motives for his actions. Inscrutable. Does not give away information for the fun of it.
I	3	Rushes in carefree fashion into new experiences, situations, emergencies. Ready to meet anything, Happy-go-lucky. Has a great appetite for life.	Avoids the strange and new. Looks at all aspects of a situation over-cautiously. Keeps clear of difficulties. Uninquiring, lacking in desire to try new things.
I	4	Likes to be in large groups. Seeks people out for the sake of company. Likes parties as often as possible. Not fond of being alone.	Does not seem to miss company of others. Goes his own way.
II	1	Does not mind when people use his property, time or energy. Generous, gives people "the benefit of the doubt" when their motives are in question. Warm-hearted.	Gets irritable, or resentful if property or other rights are trespassed on. Inclined to be "close" and grasping. Is generally surly, hard, and spiteful.
II	2	Not prone to jealousy.	Becomes readily jealous of people. Unreasonably hostile.
II	3	Gentle-tempered. Blames himself (or nobody) if things go wrong.	Goes his own way regardless of others. Blames others, not himself, whenever there is conflict or things go wrong. Headstrong. Predatory — tends to use other people for his own ends.
II	4	Generally tends to say yes when invited to cooperate. Outgoing. Ready to meet people at least half way. Finds ways of cooperating despite difficulties.	Inclined to raise objections to a project, cynical or realistic. "Cannot be done" attitude, Uninterested or unfavorable attitude to joining in. Inclined to be "difficult."

39

Factor	Replicate	[A]	[B]
III	1	Tidy, over-precise, especially over details. Drives other people to be the same. Strict, fussy, pedantic. Insists on everything being orderly. (In these respects rather "uncomfortable to live with".) Seems unable to relax. Miserly.	Rather careless of detail. Lazy. Careless over expenditures. Has no difficulty in relaxing. Enjoys ease.
III	2	Has a sense of responsibility to his parents, community, etc. Can be depended upon to be loyal to agreed standards. Trustworthy.	Does not seem to take responsibilities seriously. Undependable. Thoughtless. Refuses to accept responsibilities of his age.
III	3	Careful about principles of conduct. Guided by ideals, ethics, unselfishness. Scrupulously upright where personal desires conflict with principles.	Inclined to somewhat shady transactions. Not too careful about right and wrong where own wishes are concerned. Not particularly just, ethical, or unselfish.
III	4	Sees a job through in spite of difficulties or temptations. Strong-willed. Persisting in his motives. Painstaking and thorough.	Gives up rather easily. Led astray from main purposes by stray impulses. Slipshod — does not finish a job thoroughly.
IV	1	Rarely seems to get tired or upset. Goes on with what he is doing regardless of distractions. Rarely shows any nervousness.	Easily gets tired and overwrought. Is frequently irritable. Jumps when spoken to. Shows occasional signs of "nervousness" (e.g., fidgeting, tremor, digestive disturbances, poor memory). Constantly complains of fatigue.
IV	2	Calm, tough. "What's the fuss about?" attitude.	Worries constantly, sensitive, hurried; seems to suffer from more anxieties than other people. Slight suppressed agitation most of the time.
IV	3	Self-possessed, hard. Does not lose composure, e.g., through emotional provocation.	Easily embarrassed or put off balance in conversation. Gets confused in emergency. Blushes, shows excitability, becomes incoherent. (Not general emotionality, but momentary "nervousness.")
IV	4	Does not worry about illnesses.	Dwells on illness or hurts a great deal. Magnifies relatively trivial illnesses. Fusses a good deal over bodily symptoms.

40

Factor	Replicate	[A]	[B]
V	1	Artistically sensitive to surroundings. Fastidious, not too easily pleased.	Not showing artistic taste. Not interested in artistic subjects. Insensitive to esthetic effects.
V	2	Has wide interest and knowledge, especially in intellectual matters. Enjoys analytical, penetrating discussions in small groups.	Rather ignorant. Unreflective. Does not read much or enjoy intellectual problems. Narrow, simple, interests.
V	3	Polite and charming in social situations. Deals with people gracefully and skillfully. Refined with speech, manner, etc. Familiar with good etiquette.	Clumsy in social situations. Crude in speech, manner, etc.
V	4	Inclined to be governed by a vivid imagination. Thinks of unusual angles and aspects of a question. Sensitive to a multitude of emotional and other possibilities not realized by the average person. Intuitive, more interested in mental than material and practical aspects of a situation.	Solves questions in a logical matter-of-fact fashion which often ignores fine points or unusual possibilities. Heavily and "blindly" logical, refusing to see intangibles. More interested in material than mental aspects of a situation.

Div. 23/1, 28/4

Aeronautical Systems Division. Personnel Laboratory, Lackland Air Force Base, Texas. DEVELOPMENT OF SELF-REPORT TESTS TO MEASURE PERSONALITY FACTORS IDENTIFIED FROM PEER NOMINATIONS, by Warren T. Norman. May 1961. vi + 41 p. (Project 7717; Task 87003) (ASD-TN-61-44) (Contract AF 41(657)-269, University of Michigan)   Unclassified report

An experimental battery of personality tests was constructed as part of a project to develop personality tests appropriate for use in selection of applicants for Air Force officer training. Criteria were peer-nomination ratings previously shown to define personality factors that were predictive of Officer Effectiveness Ratings. Rational selection of testing techniques and item forms was

(over)

supplemented by information from a series of tryouts with small samples. The battery will be administered to a large sample composed of groups from which reliable peer-rating criteria can be obtained for full cross validation.